

双论域粗糙集理论与方法

杨海龙 著



科学出版社

陕西师范大学优秀研究生教材资助项目

双论域粗糙集理论与方法

杨海龙 著

科学出版社

北京

内 容 简 介

本书系统研究双论域上的粗糙集理论与方法,包括单论域粗糙集的基本理论、双论域上的粗糙集模型的变换、双论域上基于水平集的粗糙集、双论域上的概率粗糙集、双论域上的多粒度粗糙集与多粒度概率粗糙集、双论域上的多粒度决策粗糙集等理论及方法。

本书可供高等院校数学、计算机科学、信息科学、自动化、管理科学与工程等专业的高年级本科生、研究生参考,也可作为从事粗糙集理论与方法研究的科技人员的参考书。

图书在版编目(CIP)数据

双论域粗糙集理论与方法/杨海龙著. —北京:科学出版社,2016.5
ISBN 978-7-03-048270-9

I. ①双… II. ①杨… III. ①集论—研究 IV. ①O144

中国版本图书馆 CIP 数据核字(2016)第 102407 号

责任编辑:王 哲 董素芹 / 责任校对:蒋 萍

责任印制:张 倩 / 封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

文林印务有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 5 月第 一 版 开本:720×1 000 1/16

2016 年 5 月第一次印刷 印张:14 1/4

字数:285 000

定价:86.00 元

(如有印装质量问题,我社负责调换)

前 言

粗糙集理论是波兰科学院院士 Pawlak 于 1982 年提出的，其在人工智能、机器学习、数据挖掘以及决策分类等领域有着广泛的应用。本书主要介绍双论域上的粗糙集理论与方法。在内容安排上，本书主要是由作者近几年在该领域的最新研究成果构成的，为了内容上的层次性和完整性，同时也融入了国内外部分研究者在该领域里所做的一些重要工作。

全书共分 6 章。第 1 章作为预备知识介绍单论域粗糙集的基本理论；第 2 章讲述双论域上的粗糙集模型的变换，包括双论域上的分明粗糙集模型的变换、双论域上的模糊粗糙集模型的变换、双论域上的直觉模糊粗糙集模型的变换等；第 3 章分别以模糊关系、直觉模糊关系、区间值模糊关系为基点，研究双论域上基于水平集的粗糙集，并阐述提出的粗糙集模型的应用；第 4 章将概率论应用到双论域粗糙集，研究双论域上的概率粗糙集及其应用；第 5 章将多粒度思想应用到双论域粗糙集及概率粗糙集，研究双论域上的多粒度粗糙集和多粒度概率粗糙集，包含双论域上的乐观多粒度粗糙集、悲观多粒度粗糙集、加权平均多粒度概率粗糙集、乐观多粒度概率粗糙集和悲观多粒度概率粗糙集等；第 6 章以完备信息和不完备信息下的多粒度决策粗糙集为基础，利用贝叶斯理论提出并详细研究双论域上的多粒度决策粗糙集。

本书的出版得到了国家自然科学基金项目（项目编号：61473181）、陕西师范大学研究生教育教学改革研究项目（项目编号：GERP-14-39）以及作者所属数学与信息科学学院一流学科建设经费的资助。本书的顺利出版离不开学院各位领导的大力支持与帮助，在此一并表示感谢！

最后，对本书中的不足之处，恳请广大同行和读者批评指正。

杨海龙

2016 年 4 月于西安

目 录

前言

第 1 章	单论域粗糙集的基本理论	1
1.1	Pawlak 粗糙集和基于一般二元关系的粗糙集	1
1.1.1	Pawlak 粗糙集	1
1.1.2	基于一般二元关系的粗糙集	7
1.2	模糊粗糙集	11
1.3	概率粗糙集	12
1.4	决策粗糙集	16
1.5	多粒度粗糙集	17
	参考文献	24
第 2 章	双论域上的粗糙集模型的变换	26
2.1	双论域上的分明粗糙集模型的变换	26
2.1.1	双论域上的分明粗糙集模型	26
2.1.2	双论域上的分明粗糙集模型的并	28
2.1.3	双论域上的分明粗糙集模型的交	29
2.1.4	双论域上的分明粗糙集模型的逆	30
2.1.5	双论域上的分明粗糙集模型的合成	32
2.2	双论域上的模糊粗糙集模型的变换	33
2.2.1	双论域上的模糊粗糙集模型	33
2.2.2	双论域上的模糊粗糙集模型的并	34
2.2.3	双论域上的模糊粗糙集模型的交	35
2.2.4	双论域上的模糊粗糙集模型的逆	37
2.2.5	双论域上的模糊粗糙集模型的合成	39
2.3	双论域上的直觉模糊粗糙集模型的变换	41
2.3.1	双论域上的直觉模糊粗糙集模型	41
2.3.2	双论域上的直觉模糊粗糙集模型的并	42
2.3.3	双论域上的直觉模糊粗糙集模型的交	44
2.3.4	双论域上的直觉模糊粗糙集模型的逆	46
2.3.5	双论域上的直觉模糊粗糙集模型的合成	49
2.3.6	应用举例	51

参考文献	52
第3章 双论域上基于水平集的粗糙集	56
3.1 双论域上基于水平集的 1-型模糊粗糙集	56
3.1.1 双论域上基于水平集的 1-型模糊粗糙集的概念与性质	56
3.1.2 双论域上基于水平集的 1-型模糊粗糙集的推广	60
3.2 双论域上基于水平集的 2-型模糊粗糙集	63
3.2.1 双论域上基于水平集的 2-型模糊粗糙集的概念与性质	63
3.2.2 双论域上基于水平集的 2-型模糊粗糙集的推广	66
3.2.3 应用举例	68
3.3 双论域上基于水平集的 1-型直觉模糊粗糙集	69
3.3.1 双论域上基于水平集的 1-型直觉模糊粗糙集的概念与性质	69
3.3.2 双论域上基于水平集的 1-型直觉模糊粗糙集的推广	74
3.4 双论域上基于水平集的 2-型直觉模糊粗糙集	77
3.4.1 双论域上基于水平集的 2-型直觉模糊粗糙集的概念与性质	77
3.4.2 双论域上基于水平集的 2-型直觉模糊粗糙集的推广	80
3.4.3 应用举例	83
3.5 双论域上基于水平集的 1-型区间值直觉模糊粗糙集	84
3.5.1 双论域上基于水平集的 1-型区间值直觉模糊粗糙集的概念与性质	84
3.5.2 双论域上基于水平集的 1-型区间值直觉模糊粗糙集的推广	89
3.6 双论域上基于水平集的 2-型区间值直觉模糊粗糙集	92
3.6.1 双论域上基于水平集的 2-型区间值直觉模糊粗糙集的概念与性质	92
3.6.2 双论域上基于水平集的 2-型区间值直觉模糊粗糙集的推广	96
3.6.3 应用举例	98
参考文献	99
第4章 双论域上的概率粗糙集	104
4.1 双论域上基于一般二元关系的概率粗糙集	104
4.1.1 双论域上基于一般二元关系的概率粗糙集的概念与性质	104
4.1.2 应用实例	110
4.2 双论域上的模糊概率粗糙集	111
4.2.1 双论域上模糊概率粗糙集的概念与性质	111
4.2.2 应用实例	117
4.3 双论域上的直觉模糊概率粗糙集	119
4.3.1 双论域上直觉模糊概率粗糙集的概念与性质	119
4.3.2 应用实例	126
参考文献	128

第 5 章 双论域上的多粒度粗糙集与多粒度概率粗糙集	133
5.1 双论域上的多粒度粗糙集	133
5.1.1 双论域上的乐观多粒度粗糙集	133
5.1.2 双论域上的悲观多粒度粗糙集	137
5.2 双论域上的多粒度概率粗糙集	143
5.2.1 双论域上的加权平均多粒度概率粗糙集	143
5.2.2 双论域上的乐观多粒度概率粗糙集	148
5.2.3 双论域上的悲观多粒度概率粗糙集	152
参考文献	159
第 6 章 双论域上的多粒度决策粗糙集	164
6.1 完备信息下的多粒度决策粗糙集	164
6.1.1 完备信息下的加权平均多粒度决策粗糙集	164
6.1.2 完备信息下的乐观多粒度决策粗糙集	168
6.1.3 完备信息下的悲观多粒度决策粗糙集	174
6.2 不完备信息下的多粒度决策粗糙集	180
6.2.1 不完备信息下的加权平均多粒度决策粗糙集	181
6.2.2 不完备信息下的乐观多粒度决策粗糙集	184
6.2.3 不完备信息下的悲观多粒度决策粗糙集	190
6.3 双论域上基于串行二元关系的多粒度决策粗糙集	199
6.3.1 双论域上的加权平均多粒度决策粗糙集	199
6.3.2 双论域上的乐观多粒度决策粗糙集	203
6.3.3 双论域上的悲观多粒度决策粗糙集	209
参考文献	215

第 1 章 单论域粗糙集的基本理论

作为本书后续章节的预备知识,本章集中给出单论域粗糙集的一些基本理论,内容包括: Pawlak 粗糙集、基于一般二元关系的粗糙集、模糊粗糙集、概率粗糙集、决策粗糙集以及多粒度粗糙集等。

1.1 Pawlak 粗糙集和基于一般二元关系的粗糙集

1.1.1 Pawlak 粗糙集

1982 年波兰科学院院士 Pawlak 教授提出了基于等价关系的粗糙集概念,称为 Pawlak 粗糙集。下面首先给出此定义。

定义 1.1 设 U 是一个非空有限论域, R 是 U 上的一个等价关系, (U, R) 称为一个 Pawlak 近似空间, $\forall X \subseteq U$, 定义两个子集

$$\underline{R}(X) = \bigcup \{[x]_R \mid [x]_R \subseteq X\}$$

$$\bar{R}(X) = \bigcup \{[x]_R \mid [x]_R \cap X \neq \emptyset\}$$

分别称 $\underline{R}(X)$ 和 $\bar{R}(X)$ 为 X 的下近似和上近似。 $(\underline{R}(X), \bar{R}(X))$ 称为 Pawlak 粗糙集。 $\underline{R}$ 和 $\bar{R}$ 分别称为下近似算子和上近似算子。系统 $(P(U), \cap, \cup, \underline{R}, \bar{R})$ 称为粗糙集代数。

注 1.1 定义 1.1 中的下近似、上近似可等价地用下面的等式来表示, 即

$$\underline{R}(X) = \{x \in U \mid [x]_R \subseteq X\}$$

$$\bar{R}(X) = \{x \in U \mid [x]_R \cap X \neq \emptyset\}$$

基于集合 X 的下近似、上近似, X 的正域 $\text{POS}(X)$ 、负域 $\text{NEG}(X)$ 和边界域 $\text{BND}(X)$ 分别定义为

$$\text{POS}(X) = \underline{R}(X) = \{x \in U \mid [x]_R \subseteq X\}$$

$$\text{NEG}(X) = U - \bar{R}(X) = \{x \in U \mid [x]_R \cap X = \emptyset\} = \{x \in U \mid [x]_R \subseteq X^c\}$$

$$\text{BNG}(X) = \bar{R}(X) - \underline{R}(X) = (\text{POS}(X) \cup \text{NEG}(X))^c$$

$$= \{x \in U \mid [x]_R \cap X \neq \emptyset \text{ 且 } [x]_R \cap X^c \neq \emptyset\}$$

如果将 R 看成已知知识, 则 $\underline{R}(X)$ 或 $\text{POS}(X)$ 是由那些根据知识 R 判断肯定属于 X

的 U 中元素组成的集合; $\bar{R}(X)$ 是由那些根据知识 R 判断可能属于 X 的 U 中元素组成的集合; $\text{BND}(X)$ 是那些根据知识 R 既不能判断肯定属于 X 又不能判断肯定属于 X^c 的 U 中元素组成的集合; $\text{NEG}(X)$ 是那些根据知识 R 判断肯定不属于 X 的 U 中元素组成的集合。

由定义 1.1 可得下近似算子和上近似算子具有以下性质。

定理 1.1 设 (U, R) 为 Pawlak 近似空间, $\forall X, Y \subseteq U$, 则有如下结论。

- (1) $\underline{R}(X) \subseteq X \subseteq \bar{R}(X)$ 。
- (2) $\underline{R}(\emptyset) = \bar{R}(\emptyset) = \emptyset$, $\underline{R}(U) = \bar{R}(U) = U$ 。
- (3) $\underline{R}(X \cap Y) = \underline{R}(X) \cap \underline{R}(Y)$, $\bar{R}(X \cup Y) = \bar{R}(X) \cup \bar{R}(Y)$ 。
- (4) 如果 $X \subseteq Y$, 则 $\underline{R}(X) \subseteq \underline{R}(Y)$ 且 $\bar{R}(X) \subseteq \bar{R}(Y)$ 。
- (5) $\underline{R}(X \cup Y) \supseteq \underline{R}(X) \cup \underline{R}(Y)$, $\bar{R}(X \cap Y) \subseteq \bar{R}(X) \cap \bar{R}(Y)$ 。
- (6) $\underline{R}(X) = (\bar{R}(X^c))^c$, $\bar{R}(X) = (\underline{R}(X^c))^c$ 。
- (7) $\underline{R}(\underline{R}(X)) = \bar{R}(\bar{R}(X)) = \underline{R}(X)$, $\bar{R}(\bar{R}(X)) = \underline{R}(\underline{R}(X)) = \bar{R}(X)$ 。

证明 (1) $\forall x \in \underline{R}(X)$, 则 $[x]_R \subseteq X$, 而 $x \in [x]_R$, 所以 $x \in X$, 因此 $\underline{R}(X) \subseteq X$ 。

(2) 由结论 (1) 知 $\underline{R}(\emptyset) \subseteq \emptyset$, 而显然 $\emptyset \subseteq \underline{R}(\emptyset)$, 因此 $\underline{R}(\emptyset) = \emptyset$ 。

假设 $\bar{R}(\emptyset) \neq \emptyset$, 则存在 $x \in U$, 使得 $x \in \bar{R}(\emptyset)$, 即 $[x]_R \cap \emptyset \neq \emptyset$, 这与 $[x]_R \cap \emptyset = \emptyset$ 矛盾, 因此 $\bar{R}(\emptyset) = \emptyset$ 。

由结论 (1) 知 $\underline{R}(U) \subseteq U$ 。 $\forall x \in U$, 有 $[x]_R \subseteq U$, 所以 $x \in \underline{R}(X)$, 则 $U \subseteq \underline{R}(X)$, 因此 $\underline{R}(U) = U$ 。

由结论 (1) 可知 $\bar{R}(U) \supseteq U$, 又显然 $\bar{R}(U) \subseteq U$, 故 $\bar{R}(U) = U$ 。

(3) 因为 $x \in \underline{R}(X \cap Y) \Leftrightarrow [x]_R \subseteq X \cap Y$

$$\Leftrightarrow [x]_R \subseteq X \text{ 且 } [x]_R \subseteq Y$$

$$\Leftrightarrow x \in \underline{R}(X) \cap \underline{R}(Y)$$

所以 $\underline{R}(X \cap Y) = \underline{R}(X) \cap \underline{R}(Y)$ 。

因为 $x \in \bar{R}(X \cup Y) \Leftrightarrow [x]_R \cap (X \cup Y) \neq \emptyset$

$$\Leftrightarrow ([x]_R \cap X) \cup ([x]_R \cap Y) \neq \emptyset$$

$$\Leftrightarrow [x]_R \cap X \neq \emptyset \text{ 或 } [x]_R \cap Y \neq \emptyset$$

$$\Leftrightarrow x \in \bar{R}(X) \text{ 或 } x \in \bar{R}(Y)$$

$$\Leftrightarrow x \in \bar{R}(X) \cup \bar{R}(Y)$$

所以 $\bar{R}(X \cup Y) = \bar{R}(X) \cup \bar{R}(Y)$ 。

(4) 如果 $X \subseteq Y$, 则 $X \cap Y = X$, 所以 $\underline{R}(X \cap Y) = \underline{R}(X)$ 。由结论 (3) 知, $\underline{R}(X) \cap \underline{R}(Y) = \underline{R}(X)$, 因此 $\underline{R}(X) \subseteq \underline{R}(Y)$ 。

如果 $X \subseteq Y$, 则 $X \cup Y = Y$, 所以 $\bar{R}(X \cup Y) = \bar{R}(Y)$ 。由结论(3)知, $\bar{R}(X) \cup \bar{R}(Y) = \bar{R}(Y)$, 因此 $\bar{R}(X) \subseteq \bar{R}(Y)$ 。

(5) 因为 $X \subseteq X \cup Y$, $Y \subseteq X \cup Y$, 所以由结论(4)知, $\underline{R}(X) \subseteq \underline{R}(X \cup Y)$ 且 $\underline{R}(Y) \subseteq \underline{R}(X \cup Y)$, 因此 $\underline{R}(X) \cup \underline{R}(Y) \subseteq \underline{R}(X \cup Y)$ 。

因为 $X \cap Y \subseteq X$, $X \cap Y \subseteq Y$, 所以由结论(4)知, $\bar{R}(X \cap Y) \subseteq \bar{R}(X)$ 且 $\bar{R}(X \cap Y) \subseteq \bar{R}(Y)$, 故 $\bar{R}(X \cap Y) \subseteq \bar{R}(X) \cap \bar{R}(Y)$ 。

(6) 因为 $x \in \underline{R}(X) \Leftrightarrow [x]_R \subseteq X$

$$\Leftrightarrow [x]_R \cap X^c = \emptyset$$

$$\Leftrightarrow x \notin \bar{R}(X^c)$$

$$\Leftrightarrow x \in (\bar{R}(X^c))^c$$

所以 $\underline{R}(X) = (\bar{R}(X^c))^c$ 。

由 $\underline{R}(X) = (\bar{R}(X^c))^c$ 可知, $(\underline{R}(X^c))^c = ((\bar{R}((X^c))^c)^c)^c = \bar{R}(X)$ 。

(7) 由结论(1)知, $\underline{R}(\underline{R}(X)) \subseteq \underline{R}(X)$ 。 $\forall x \in \underline{R}(X)$, 有 $[x]_R \subseteq X$, 因此 $\underline{R}([x]_R) \subseteq \underline{R}(X)$ 。显然 $\underline{R}([x]_R) = [x]_R$, 于是 $[x]_R \subseteq \underline{R}(X)$, 所以 $x \in \underline{R}(\underline{R}(X))$, 从而 $\underline{R}(X) \subseteq \underline{R}(\underline{R}(X))$ 。故 $\underline{R}(X) = \underline{R}(\underline{R}(X))$ 。

由结论(1)知, $\underline{R}(X) \supseteq \bar{R}(\underline{R}(X))$ 。 $\forall x \in \bar{R}(\underline{R}(X))$, 有 $[x]_R \cap \underline{R}(X) \neq \emptyset$, 即存在 $y \in [x]_R$ 且 $y \in \underline{R}(X)$, 所以 $[y]_R \subseteq X$, 而 $[y]_R = [x]_R$, 则 $[x]_R \subseteq X$, 即 $x \in \underline{R}(X)$, 因此 $\bar{R}(X) \supseteq \bar{R}(\underline{R}(X))$ 。故 $\bar{R}(\underline{R}(X)) = \underline{R}(X)$ 。

由结论(1)知, $\bar{R}(\bar{R}(X)) \subseteq \bar{R}(X)$ 。 $\forall x \in \bar{R}(\bar{R}(X))$, 有 $[x]_R \cap \bar{R}(X) \neq \emptyset$, 则存在 $y \in [x]_R$ 且 $y \in \bar{R}(X)$, 所以 $[y]_R \cap X \neq \emptyset$ 。而 $[x]_R = [y]_R$, 所以 $[x]_R \cap X \neq \emptyset$, 即 $x \in \bar{R}(X)$, 因此 $\bar{R}(X) \supseteq \bar{R}(\bar{R}(X))$ 。故 $\bar{R}(\bar{R}(X)) = \bar{R}(X)$ 。

由结论(1)知, $\underline{R}(\bar{R}(X)) \subseteq \bar{R}(X)$ 。 $\forall x \in \bar{R}(X)$, 有 $[x]_R \cap X \neq \emptyset$, 所以 $[x]_R \subseteq \bar{R}(X)$ 。这是因为 $\forall y \in [x]_R$, 有 $[y]_R \cap X = [x]_R \cap X \neq \emptyset$, 则 $y \in \bar{R}(X)$ 。所以 $x \in \underline{R}(\bar{R}(X))$, 从而 $\underline{R}(\bar{R}(X)) \supseteq \bar{R}(X)$ 。故 $\underline{R}(\bar{R}(X)) = \bar{R}(X)$ 。

在不致混淆的情况下, $[x]_R$ 可以简写为 $[x]$ 。

集合的不精确性是由于边界域的存在而引起的, 边界域越大, 其精确性越低。为了更准确地表达这一点, 下面引入近似精度的概念。由等价关系 R 定义的集合 $X (X \neq \emptyset)$ 的近似精度为

$$\alpha_R(X) = \frac{|\underline{R}(X)|}{|\bar{R}(X)|}$$

其中, $|\cdot|$ 表示集合的基数(势)。

显然 $0 \leq \alpha_R(X) \leq 1$, 且 $\alpha_R(X) = 1$ 当且仅当 X 的边界域为空。

利用精度 $\alpha_R(X)$ 可以定义集合 X 的不精确程度。 X 的近似粗糙度定义为

$$\rho_R(X) = 1 - \alpha_R(X)$$

下面给出近似分类的概念。

定义 1.2 设 $F = \{X_1, X_2, \dots, X_n\}$ 是论域 U 的一个分类或划分, 它可能由一个专家为解决一个分类问题所给出。 F 的 R 下近似和上近似分别定义为

$$\underline{R}F = \{\underline{R}(X_1), \underline{R}(X_2), \dots, \underline{R}(X_n)\}$$

$$\bar{R}F = \{\bar{R}(X_1), \bar{R}(X_2), \dots, \bar{R}(X_n)\}$$

定义 1.3 (1) 称 $\alpha_R(F) = \frac{\sum_{i=1}^n |\underline{R}(X_i)|}{\sum_{i=1}^n |\bar{R}(X_i)|}$ 为 F 的近似分类精度。

(2) 称 $\gamma_R(X) = \frac{\sum_{i=1}^n |\underline{R}(X_i)|}{|U|}$ 为 F 的近似分类质量。

近似分类精度描述的是当使用知识 R 对对象分类时, 可能的决策中正确决策的百分比; 分类的质量表示的是应用知识 R 能确切地划入 F 类的对象的百分比。

现在介绍集合的三种近似相等的概念。

定义 1.4 设 (U, R) 是一个 Pawlak 近似空间, $\forall X, Y \subseteq U$, 有如下结论。

(1) 若 $\underline{R}(X) = \underline{R}(Y)$, 则称集合 X 和 Y 是下粗相等, 记为 $X \preceq Y$ 。

(2) 若 $\bar{R}(X) = \bar{R}(Y)$, 则称集合 X 和 Y 是上粗相等, 记为 $X \succeq Y$ 。

(3) 若 $X \preceq Y$ 且 $X \succeq Y$, 则称集合 X 和 Y 是粗相等, 记为 $X \approx Y$ 。

易知, $\preceq$ 、 $\succeq$ 和 $\approx$ 是等价关系。由定义 1.4 知它们具有下列基本性质。

定理 1.2 设 (U, R) 是一个 Pawlak 近似空间, $X, X', Y, Y' \subseteq U$, 有如下结论。

(1) $X \preceq Y$, 当且仅当 $(X \cap Y) \preceq X$ 且 $(X \cap Y) \preceq Y$ 。

(2) $X \succeq Y$, 当且仅当 $(X \cup Y) \succeq X$ 且 $(X \cup Y) \succeq Y$ 。

(3) 若 $X \preceq X'$ 且 $Y \preceq Y'$, 则 $X \cup Y \preceq X' \cup Y'$ 。

(4) 若 $X \preceq X'$ 且 $Y \preceq Y'$, 则 $X \cap Y \preceq X' \cap Y'$ 。

(5) 若 $X \preceq Y$, 则 $X \cup Y^c \preceq U$ 。

(6) 若 $X \preceq Y$, 则 $X \cap Y^c \preceq \emptyset$ 。

(7) 若 $X \subseteq Y$ 且 $Y \preceq \emptyset$, 则 $X \preceq \emptyset$ 。

(8) 若 $X \subseteq Y$ 且 $X \preceq U$, 则 $Y \preceq U$ 。

(9) $X \preceq Y$ 当且仅当 $X^c \preceq Y^c$ 。

(10) 若 $X \preceq \emptyset$ 或 $Y \preceq \emptyset$, 则 $X \cap Y \preceq \emptyset$ 。

(11) 若 $X \approx U$ 或 $Y \approx U$ ，则 $X \cup Y \approx U$ 。

证明 由定义 1.4 易证。

可以用集合的粗糙相等来表示下近似和上近似。

定理 1.3 设 (U, R) 是一个 Pawlak 近似空间，则有如下结论。

(1) $\underline{R}(X)$ 是 U 中所有满足 $X \approx Y$ 的集合 Y 的交。

(2) $\overline{R}(X)$ 是 U 中所有满足 $X \approx Y$ 的集合 Y 的并。

证明 利用集合的包含关系易证。

下面给出信息表的概念。

四元组 $I = (U, AT, V, \{f_a | a \in A\})$ 称为一个信息表，其中 U 为对象的非空有限集合，称为论域； AT 为属性的非空有限集合； $V = \bigcup_{a \in A} V_a$ ， V_a 是属性 a 的值域； $f_a: U \rightarrow V_a$ 是一个信息函数， $\forall a \in A$ ， f_a 为每个对象赋予一个属性 a 的值，即 $x \in U$ ， $f_a(x) \in V_a$ 。

信息表的数据以关系表的形式表示，关系表的行对应要研究的对象，列对应对象的属性，对象的信息是通过对象的各属性值来表达的。

$\forall A \subseteq AT$ ，由 A 可诱导出一个不可区分关系（是一个等价关系） $IND(A)$ 如下

$$IND(A) = \{(x, y) \in U \times U | \forall a \in A, f_a(x) = f_a(y)\}$$

易见，一个属性对应一个等价关系，从而一个信息表可以得到一簇等价关系。

例 1.1 设 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ 是患者的集合， $AT = \{a_1, a_2, a_3\}$ 是属性集，其中 a_1 表示“头痛”， a_2 表示“肌肉痛”， a_3 表示“体温”。

设有一个关于某些患者的信息表，如表 1.1 所示。

表 1.1 患者的信息表

U	a_1	a_2	a_3
x_1	是	是	正常
x_2	是	是	高
x_3	是	是	很高
x_4	否	是	正常
x_5	否	否	高
x_6	否	是	很高

取 $A = \{a_1, a_2\}$ ，则由 A 诱导出一个不可区分关系，如表 1.2 所示。

表 1.2 不可区分关系 $IND(A)$

$IND(A)$	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1	1	1	0	0	0
x_2	1	1	1	0	0	0
x_3	1	1	1	0	0	0

续表

IND(A)	x_1	x_2	x_3	x_4	x_5	x_6
x_4	0	0	0	1	0	1
x_5	0	0	0	0	1	0
x_6	0	0	0	1	0	1

即 $U/IND(A) = \{\{x_1, x_2, x_3\}, \{x_4, x_6\}, \{x_5\}\}$ 。取 $X = \{x_2, x_4, x_6\}$ ，由定义 1.1 可得

$$\underline{A}X = \{x_4, x_6\}, \quad \overline{A}X = \{x_1, x_2, x_3, x_4, x_6\}$$

于是

$$POS(X) = \{x_4, x_6\}, \quad NEG(X) = \{x_5\}, \quad BND(X) = \{x_1, x_2, x_3\}$$

上面提及的信息表的每个对象的所有属性值均为已知的。如果信息表中某个对象的属性值是未知或部分已知的，则称这种信息表为不完备信息表。对于完全不知道的情形，称这种属性值为空值。空值一般有三种情形：①不存在型空值，即无法填入的值或对象在该属性上无法取值；②存在型空值，即对象在该属性上取值是存在的，但暂时无法知道，此时往往将其可能取值的全体（即该属性的值域）作为该对象的属性取值；③占位型空值，即无法确定是不存在型还是存在型空值。本书所指的空值是存在型空值，用“*”表示这个值未知。对于部分知道的情形，即肯定不取某值，则它有可能取该属性值域中的其他值的任何一个。在不致混淆的情况下，不完备信息表仍用记号 $I = (U, AT, V, \{f_a \mid a \in A\})$ 表示。

对于一个不完备信息表 $I = (U, AT, V, \{f_a \mid a \in A\})$ ， $A \subseteq AT$ ，由 A 可诱导如下一个二元关系，记为 R_A

$$R_A = \{(x, y) \in U \times U \mid \forall a \in A, f(x, a) = f(y, a) \text{ 或 } f(x, a) = * \text{ 或 } f(y, a) = *\}$$

易见， R_A 是一个容差关系，即自反的且对称的。

例 1.2 设 $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ ， $AT = \{a_1, a_2, a_3, a_4\}$ ， $I = (U, AT, V, f)$ 是一个不完备信息表，如表 1.3 所示。

表 1.3 不完备信息 I

U	a_1	a_2	a_3	a_4
x_1	*	2	*	3
x_2	1	3	3	2
x_3	2	2	3	1
x_4	2	3	1	*
x_5	3	*	1	1
x_6	2	1	*	3

取 $A = \{a_2, a_3\}$ ，则由 A 诱导出一个二元关系，如表 1.4 所示。

表 1.4 容差关系 R_A

R_A	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1	0	1	0	1	0
x_2	0	1	0	0	0	0
x_3	1	0	1	0	0	0
x_4	0	0	0	1	1	0
x_5	1	0	0	1	1	1
x_6	0	0	0	0	1	1

下面给出决策表和不完备决策表的概念。

设 $I = (U, AT, V, \{f_a \mid a \in A\})$ 是一个信息表, $AT = C \cup D$, $C \cap D = \emptyset$, C 称为条件属性集, D 称为决策属性集。具有条件属性和决策属性的信息表称为决策表。具有条件属性和决策属性的不完备信息表称为不完备决策表。

例 1.3 设 $U = \{x_1, x_2, x_3, x_4, x_5\}$ 表示患者的集合, $C = \{a_1, a_2, a_3\}$, 其中 a_1 表示“头痛”, a_2 表示“肌肉痛”, a_3 表示“体温”, $D = \{d\}$, d 表示“流感”。 $I = (U, C \cup D, V, \{f_a \mid a \in A\})$ 是一个决策表, 如表 1.5 所示。

表 1.5 决策表 I

U	a_1	a_2	a_3	d
x_1	是	是	正常	否
x_2	是	是	高	是
x_3	是	是	很高	是
x_4	否	是	正常	否
x_5	否	否	高	否

1.1.2 基于一般二元关系的粗糙集

定义 1.5 设 R 是 U 上的二元关系, 对于 $x, y \in U$, 若 xRy 即 $(x, y) \in R$, 则称 x 是 y 的前继, y 是 x 的后继, 记

$$R_s(x) = \{y \in U \mid (x, y) \in R\}$$

$$R_p(x) = \{y \in U \mid (y, x) \in R\}$$

$$R_{p \wedge s}(x) = \{y \in U \mid (x, y) \in R, (y, x) \in R\} = R_p(x) \cap R_s(x)$$

$$R_{p \vee s}(x) = \{y \in U \mid (x, y) \in R \text{ 或 } (y, x) \in R\} = R_s(x) \cup R_p(x)$$

分别称为 x 的后继邻域、前继邻域、下继邻域和上继邻域。

显然

$$R_{p \wedge s}(x) \subseteq R_p(x) \subseteq R_{p \vee s}(x)$$

$$R_{p \wedge s}(x) \subseteq R_s(x) \subseteq R_{p \vee s}(x)$$

$$R_p(x) = R_s^{-1}(x)$$

定义 1.6 设 R 是 U 上的二元关系。

(1) 若 $\forall x \in U, \exists y \in U$, 使得 $y \in R_s(x)$, 则称 R 是串行的。

(2) 若 $\forall x \in U, \exists y \in U$, 使得 $x \in R_s(y)$, 则称 R 是逆串行的。

(3) 若 $\forall x \in U, x \in R_s(x)$, 即 $R_s(U) = U$, 其中 $R_s(U) = \bigcup_{x \in U} R_s(x)$, 则称 R 是自反的。

(4) 若 $\forall x, y \in U, x \in R_s(y) \Leftrightarrow y \in R_s(x)$, 则称 R 是对称的。

(5) 若 $\forall x, y, z \in U, y \in R_s(x)$ 和 $z \in R_s(y)$ 蕴涵 $z \in R_s(x)$, 则称 R 是传递的。

(6) 若 $\forall x, y, z \in U, y \in R_s(x)$ 和 $z \in R_s(x)$ 蕴涵 $z \in R_s(y)$, 则称 R 是欧几里得的 (Euclidean)。

定义 1.7 设 U 是有限非空论域, $R \subseteq U \times U$ 为 U 上的一个一般二元关系, 称 (U, R) 是广义近似空间。 $\forall X \subseteq U$, X 关于近似空间 (U, R) 的下近似 $\underline{\text{apr}}(X)$ 和上近似 $\overline{\text{apr}}(X)$ 分别定义为

$$\underline{\text{apr}}(X) = \{x \in U \mid R_s(x) \subseteq X\}$$

$$\overline{\text{apr}}(X) = \{x \in U \mid R_s(x) \cap X \neq \emptyset\}$$

称 $(\underline{\text{apr}}(X), \overline{\text{apr}}(X))$ 为基于一般二元关系的粗糙集。

X 关于广义近似空间 (U, R) 的正域 $\text{POS}(X)$ 、负域 $\text{NEG}(X)$ 和边界域 $\text{BND}(X)$ 分别定义为

$$\text{POS}(X) = \underline{\text{apr}}(X)$$

$$\text{NEG}(X) = U - \overline{\text{apr}}(X) = \{x \in U \mid R_s(x) \cap X = \emptyset\}$$

$$\text{BND}(X) = \overline{\text{apr}}(X) - \underline{\text{apr}}(X)$$

注 1.2 若 R 是 U 上的一个等价关系, 则定义 1.6 中的粗糙集就退化为 Pawlak 粗糙集。

显然, 近似算子 $\underline{\text{apr}}$ 和 $\overline{\text{apr}}$ 是 $P(U)$ 上的算子, 称系统 $(P(U), \cap, \cup, ^c, \underline{\text{apr}}, \overline{\text{apr}})$ 是广义粗糙集代数。

定理 1.4 下近似和上近似算子具有下列性质。

(1) $\underline{\text{apr}}(X) = (\overline{\text{apr}}(X^c))^c$, $\overline{\text{apr}}(X) = (\underline{\text{apr}}(X^c))^c$ 。

(2) $\underline{\text{apr}}(U) = U$, $\overline{\text{apr}}(\emptyset) = \emptyset$ 。

(3) $\underline{\text{apr}}(X \cap Y) = \underline{\text{apr}}(X) \cap \underline{\text{apr}}(Y)$, $\overline{\text{apr}}(X \cup Y) = \overline{\text{apr}}(X) \cup \overline{\text{apr}}(Y)$ 。

(4) 若 $X \subseteq Y$, 则 $\underline{\text{apr}}(X) \subseteq \underline{\text{apr}}(Y)$, $\overline{\text{apr}}(X) \subseteq \overline{\text{apr}}(Y)$ 。

(5) $\underline{\text{apr}}(X \cup Y) \supseteq \underline{\text{apr}}(X) \cup \underline{\text{apr}}(Y)$, $\overline{\text{apr}}(X \cap Y) \subseteq \overline{\text{apr}}(X) \cap \overline{\text{apr}}(Y)$ 。

证明 (1) 因为

$$\begin{aligned} x \in \underline{\text{apr}}(X) &\Leftrightarrow R_s(x) \subseteq X \\ &\Leftrightarrow R_s(x) \cap X^c = \emptyset \\ &\Leftrightarrow x \in (\overline{\text{apr}}(X^c))^c \end{aligned}$$

所以 $\underline{\text{apr}}(X) = (\overline{\text{apr}}(X^c))^c$ 。同理, $\overline{\text{apr}}(X) = (\underline{\text{apr}}(X^c))^c$ 。

(2) 显然成立。

(3) 因为

$$\begin{aligned} x \in \underline{\text{apr}}(X \cap Y) &\Leftrightarrow R_s(x) \subseteq X \cap Y \\ &\Leftrightarrow R_s(x) \subseteq X \text{ 且 } R_s(x) \subseteq Y \\ &\Leftrightarrow x \in \underline{\text{apr}}(X) \text{ 且 } x \in \underline{\text{apr}}(Y) \\ &\Leftrightarrow x \in \underline{\text{apr}}(X) \cap \underline{\text{apr}}(Y) \end{aligned}$$

所以 $\underline{\text{apr}}(X \cap Y) = \underline{\text{apr}}(X) \cap \underline{\text{apr}}(Y)$ 。

同理 $\overline{\text{apr}}(X \cup Y) = \overline{\text{apr}}(X) \cup \overline{\text{apr}}(Y)$ 。

(4) 由定义 1.7 直接得到。

(5) 利用性质 (4) 可得。

注 1.3 定理 1.4 的性质 (5) 一般情况下等式不成立, 如下面的例 1.4。

例 1.4 设 $U = \{a, b, c\}$, R 是 U 上的一个二元关系, 如表 1.6 所示。

表 1.6 例 1.4 二元关系 R

R	a	b	c
a	1	1	1
b	0	1	0
c	0	1	0

取 $X = \{a\}$, $Y = \{b, c\}$, 则通过计算可得

$$\underline{\text{apr}}(X) \cup \underline{\text{apr}}(Y) = \{c\} \neq U = \underline{\text{apr}}(X \cup Y)$$

再取 $A = \{a, c\}$, $B = \{b, c\}$, 则可得

$$\overline{\text{apr}}(A \cap B) = \overline{\text{apr}}(\{c\}) = \emptyset \neq \{a, b\} = \overline{\text{apr}}(A) \cap \overline{\text{apr}}(B)$$

近似算子 $\underline{\text{apr}}$ 和 $\overline{\text{apr}}$ 还具有其他若干性质, 此处不详细列举。

Pawlak 粗糙集的近似算子的定义有两种主要形式。

(i) 基于粒的定义 $\underline{R}(X)=\bigcup\{[x] \mid [x] \subseteq X\}$, $\overline{R}(X)=\bigcup\{[x] \mid [x] \cap X \neq \emptyset\}$ 。

(ii) 基于元素的定义 $\underline{R}(X)=\{x \in U \mid [x] \subseteq X\}$, $\overline{R}(X)=\{x \in U \mid [x] \cap X \neq \emptyset\}$ 。

以上两种定义是等价的, 且每一对下近似和上近似是对偶的。当 R 是 U 上的一般二元关系时, 用 $R_s(x)$ 代替上述定义中的 $[x]$, 则有如下形式。

(I) $\underline{\text{apr}}(X)=\bigcup\{R_s(x) \mid R_s(x) \subseteq X\}$, $\overline{\text{apr}}(X)=\bigcup\{R_s(x) \mid R_s(x) \cap X \neq \emptyset\}$ 。

(II) $\underline{\text{apr}}(X)=\{x \in U \mid R_s(x) \subseteq X\}$, $\overline{\text{apr}}(X)=\{x \in U \mid R_s(x) \cap X \neq \emptyset\}$ 。

一般情况下, (I) 和 (II) 不等价。由定理 1.4 的性质 (1) 知, (II) 中的下近似和上近似是对偶的, 但 (I) 中的下近似与上近似不是对偶的, 见下面的例子。

例 1.5 设 $U=\{a,b,c\}$, R 是 U 上的一个二元关系, 如表 1.7 所示。

表 1.7 例 1.5 二元关系 R

R	a	b	c
a	1	1	0
b	0	0	1
c	0	1	0

此时, $R_s(a)=\{a,b\}$, $R_s(b)=\{c\}$, $R_s(c)=\{b\}$ 。令 $X=\{a,c\}$, 则

$$(\underline{\text{apr}}(X^c))^c = \{b\}^c = \{a,c\} \neq \{a,b,c\} = \overline{\text{apr}}(X)$$

近似算子的对偶性质很重要, 能简化计算, 为解决 (I) 中定义的近似算子不是对偶的缺陷, 有以下两种推广方式。

$$\begin{aligned} \text{(II') } \underline{\text{apr}}'(X) &= \bigcup\{R_s(x) \mid R_s(x) \subseteq X\} \\ &= \{x \in U \mid \exists y \in U, \text{s.t.}, x \in R_s(y) \text{ 且 } R_s(y) \subseteq X\} \end{aligned}$$

$$\begin{aligned} \overline{\text{apr}}'(X) &= (\underline{\text{apr}}'(X^c))^c \\ &= U - \{x \in U \mid \exists y \in U, \text{s.t.}, x \in R_s(y) \text{ 且 } R_s(y) \subseteq X^c\} \\ &= \{x \in U \mid \forall y \in U, x \in R_s(y) \text{ 且 } R_s(y) \cap X \neq \emptyset\} \cup (R_s(U))^c \end{aligned}$$

$$\begin{aligned} \text{(II'') } \underline{\text{apr}}''(X) &= (\overline{\text{apr}}''(X^c))^c \\ &= U - \{x \in U \mid \exists y \in U, \text{s.t.}, x \in R_s(y) \text{ 且 } R_s(y) \cap X^c \neq \emptyset\} \\ &= \{x \in U \mid \forall y \in U, x \in R_s(y) \text{ 且 } R_s(y) \subseteq X\} \cup (R_s(U))^c \\ \overline{\text{apr}}''(X) &= \bigcup\{R_s(x) \mid R_s(x) \cap X \neq \emptyset\} \\ &= \{x \in U \mid \exists y \in U, \text{s.t.}, x \in R_s(y) \text{ 且 } R_s(y) \cap X \neq \emptyset\} \end{aligned}$$

值得指出的是定义 1.7 未考虑 $R_s(x)$ 与 X 的重叠部分的定量信息。根据定义 1.6,